DOCUMENT RESUME

ED 396 003 TM 025 343

AUTHOR Littlefield, John; Sarnoff, Ron
TITLE Rater Communication Goals in Performance Appraisals.
PUB DATE Apr 96
NOTE 7p.; Paper presented at the Annual Meeting of the American Educational Research Association (New York, NY, April 8-12, 1996).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Communication (Thought Transfer); *Evaluation Methods; *Evaluators; *Lawyers; *Performance Based Assessment; Personnel Evaluation; *Rating Scales; Supervisors
IDENTIFIERS *Goal Directed Behavior

ABSTRACT

Goal-directed performance appraisal (PA) theory includes four components: rating context, performance judgment, performance rating, and evaluation. This study focuses on the components of rating context and performance rating. For the study, the rating context was a large civil service organization that must produce documentation for attorney promotion decisions. Performance ratings were written documentation of the message a rater wished to convey to audiences who read the rating form. Whether the raters in this context communicated their PA judgments using the broad categories of attorney performance defined by the organization's PA form (case analysis and preparation skills, advocacy and communication skills, and role attitude, work habits, and leadership skills) was studied for 142 attorneys for one year and 174 for the second year. Results supported the assertion that performance appraisals by the attorney supervisors could be accurately summarized in the broad performance categories. Rater goal-directed PA theory provided a framework for interpreting these research results. (Contains 2 tables and 14 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
JOHN LITTLEFIELD
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

ED 396 003

Rater Communication Goals in Performance Appraisals

John Littlefield and Ron Sarnoff

BEST COPY AVAILABLE

Paper presented at the Annual Meeting of the American Educational Research Association

New York, New York, April, 1996

TM025343

Rater Communication Goals in Performance Appraisals

John Littlefield and Ron Sarnoff

University of Texas Health Science Center at San Antonio
Los Angeles Public Defender's Office

Theoretical Framework

The measurement roots of performance appraisals (PA) are in psycho physical scaling where mathematical relationships have been demonstrated between the intensity of a physical stimulus (e.g., a 100 decibel noise) and its perceived intensity by a human judge (Stevens, 1962). Like psycho physical scaling, PA requires a quantitative judgment regarding the perceived level of a stimulus (e.g., the quality of another person's performance). Unfortunately, there is no physical scale (e.g., decibels) to independently measure the level of a PA stimulus. Videotapes of a task being performed (e.g., a physician interviewing a patient) have been used to provide a standardized stimulus for raters to assess (Van der Vleuten et. al., 1989). However, PA most often involves judging performance in on-the-job settings where ratee tasks are numerous and the opportunities for the rater to observe are variable. In the 1960's and 70's, researchers sought an ideal PA rating format to structure rater judgments. This proved fruitless and in 1980 a comprehensive review of PA research recommended abandoning the search for ideal rating scales and focusing instead on the cognitive processes used by raters in judging performance (Landy & Farr, 1980).

Recent theoretical work regarding PA goes beyond the psycho physical scaling orientation in which job performance was measured by summing marks on a ratng form to generate a numerical score. PA is viewed as a social and communication process in which raters are goal-directed when they assign ratings (Murphy & Cleveland, 1995). Goal-directed rater PA theory includes four components: *rating context* - the organizational environment and values, *performance judgment* - the rater's private evaluation of ratee performance, *performance rating* - the numbers and written comments marked on a rating form, and *evaluation* - the way ratings are used by an organization to make personnel decisions. Goal-directed rater PA includes the psycho physical scaling orientation (i.e., the performance rating), but also recognizes the influence of the other three components.

This study focuses on two components of the goal-directed rater PA theory: *rating context* and *performance rating*. The *rating context* is a large civil service organization that must produce documentation for attorney promotion decisions. The organization has defined three broad categories of attorney performance: 1. *Case analysis and preparation skills,* 2. *Advocacy and communication skills, and* 3. *Role attitude, work habits and leadership skills.* These three categories provide a framework for raters to use in judging ratee performance and communicating those judgments to the organization. Each broad performance category is further defined by three to seven detailed performance dimensions to aid in making the categorical judgments. *Performance ratings* are written documentation of a message the rater wishes to convey to the audiences who will read the rating form. The performance ratings in this study are viewed as communications from the attorney supervisors to organizational executives and also to the individual attorneys whose performance has been evaluated.

From a test validity perspective, it is important that ratings on detailed performance dimensions can be summarized by the three broad performance categories defined on the PA form. Messick (1995) identifies a *structural* aspect of validity that appraises the extent to which the internal structure of the assessment is consistent with the structure of the performance domain.
In this context, the attorney performance domain has been defined as consisting of three broad categories each supported by detailed performance dimensions. Ratings on the detailed dimensions

should be internally consistent with the three broad categories. Research in medical education has shown that raters do not make independent decisions on detailed performance dimensions. Instead, their ratings on forms with 10 or more performance dimensions can be summarized by two factors: knowledge/problem solving and interpersonal skills (Ramsey et. al., 1993; Maxim & Dielman, 1987). These two factors provide a structure for the physician performance domain as viewed by the raters.

The primary goal of this study is to determine whether raters in this context communicate their PA judgments using the three broad categories of attorney performance as defined by the organization's PA form. If they do, then the ratings on numerous detailed performance dimensions can be summarized by the three broad categories of attorney performance and the PA system has validity in its *structural* aspect. If their ratings on detailed performance dimensions can be represented by one or two broad categories then the attorney performance domain and its associated PA form should be revised to reflect those categories.

Methods and Data Source

Ratees consisted of all Deputy Public Defender III attorneys who applied for promotion in the Law Offices of the Los Angeles County Public Defender. These attorneys represent the defense in trials of difficult criminal cases and are evaluated annually by their supervisors. During the first year of the study, the professional performance of 142 attorneys was appraised by 18 supervisors. During the second year of the study, performance of 174 attorneys was appraised by 18 supervisors. Some individual ratees were included in both years of the study. The exact number of ratees who were rated in both years cannot be determined because names were deleted from the research data files to ensure confidentiality of PA information.

Those Deputy Public Defender III attorneys who apply for promotion receive their supervisor's appraisal of 15 skills deemed important for them to advance to Deputy Public Defender IV. The 15 performance dimensions are grouped into three broad categories: *Case Analysis and Preparation* (seven dimensions), *Advocacy and Communication* (five dimensions), and *Role Attitude, Work Habits and Leadership* (three dimensions). The rater's task is to mark an *overall* rating for each of the three broad categories using a numerical scale of 1 to 12. The three *overall* ratings are used to make administrative decisions. The three *overall* ratings are not an arithmetic average of their related component dimensions, but instead, according to the rater instruction manual, are ... "comprised of the interaction of the components as applied to each individual candidate." The rater also marks the 15 performance dimensions using the same scale of 1 to 12 in order to provide feedback to ratees. Data analyses were based on the 15 performance scores for each attorney in each year of the study, but did not include the three *overall* ratings.

Principal components were calculated on 15 scores for each ratee in year one and also in year two of the study (SAS Inc., 1989). A scree test (Cattell, 1966) was used to identify factors whose eigenvalues were substantially larger than the remaining eigenvalues. Then the selected factors were rotated to a Varimax criterion.

Results

The principal components analysis produced five factors each year that accounted for over 90% of the variance in the two original data matrices as can be seen in Table 1

Table 1 - Principal Components Analysis of 15 Performance Dimensions for Year One / Year Two

	Component 1	Component 2	Component 3	Component 4	Component 5
Eigenvalue	11.8 / 11.5	.96 / .88	.47 / .73	.29 / .30	.27 / .27
Cumulative % Variance	78.7 / 76.5	85.1 / 82.4	88.2 / 87.2	90.2 / 89.3	91.9 / 91.0

In both years, the first principal component's eigenvalues were very large (11.8 and 11.5) followed by two moderately small eigenvalues then two very small eigenvalues. The Scree Test identified three eigenvalues each year that were substantially larger than the two remaining values (see Table 1). The three corresponding principal components were rotated to a Varimax criterion as shown in Table 2.

Table 2 - Rotated Factor Loadings of Three Principal Components for Year One and Year Two

Performance Dimensions by Broad Category	Year 1			Year 2		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
Case Analysis and Preparation Skills						
1. Legal research and writing	.82	.33	.31	.80	.35	.34
2. Preparation for sentencing	.82	.37	.27	.78	.43	.27
3. Recognition and analysis of legal issues	.82	.39	.29	.79	.38	.30
4. Use of experts	.81	.38	.27	.78	.45	.30
5. Effective use of investigators and paralegals	.79	.33	.39	.78	.44	.31
6. Preparation of witnesses	.78	.35	.37	.75	.42	.36
7. Organizational skills	.75	.38	.39	.75	.32	.39
Advocacy and Communication Skills						
1. Courtroom presentation	.43	.76	.37	.38	.78	.31
2. Professional relations	.39	.73	.45	.35	.79	.40
3. Experience	.50	.70	.33	.51	.74	.21
4. Case negotiations	.49	.70	.42	.45	.74	.36
5. Client relations	.43	.65	.47	.47	.72	.32
Role Attitude, Work Habits and Leadership Skills						
1. Role attitude as a defense attorney	.29	.36	.82	.27	.30	.85
2. Leadership	.37	.41	.78	.38	.44	.72
3. Work habits	.48	.37	.73	.43	.26	.81

BEST COPY AVAILABLE

Discussion

Principal component 1 is by far the most representative single summary of the 15 dimension correlation matrix (77 - 79% of total variance); however, it is difficult to interpret except as a weighted sum of ratings on 15 performance dimensions. The Scree Test looks for discontintuities in the size of successive eigenvalues in deciding where to "draw the line" regarding how many principal components to be included in the rotation. We chose to include the first three principal components because in both data sets, eigenvalues for components four and five were small (relative to the other eigenvalues) and of similar numerical size. This was admittedly a somewhat subjective decision.

Following the decision to rotate the first three principal components, the Varimax rotation produced factor alignments that correspond to the three broad performance categories defined by the PA form: *Case Analysis and Preparation , Advocacy and Communication*, and *Role Attitude, Work Habits and Leadership* (see Table 2). Traditional psycho physically oriented PA theory would argue that the three factors in this study represent *logical error,* the tendency of raters to give similar ratings to performance dimensions perceived as logically-related (Guilford, 1936). In contrast, rater goal-directed PA theory would describe these results as supporting the organization's goal of defining the attorney performance domain as consisting of three broad categories. The *rating context* for this study is a civil service organization that must provide "objective documentation" for employee promotion decisions. The *performance ratings* are judgments by attorney supervisors whose job responsibilities include making employee promotion decisions. The results of the factor rotation support the assertion that raters in this context communicate their PA judgments using the three broad categories of attorney performance as defined by the organization's PA form. From a test validity perspective, there is support for the structural aspect of validity in this PA system.

The small number of factors in this study and in the medical education studies (Ramsey et. al., 1993; Maxim & Dielman, 1987) could be explained as an indication of expert reasoning by the raters. Psychological characteristics of expert reasoning have been described by Glaser and Chi (1988). They note that experts: 1. perceive large meaningful patterns in their domain of expertise, 2. have strong self-monitoring skills, 3. analyze a problem qualitatively and build a mental representation that defines the situation, and 4. cognitively represent a problem using a small number of principle-based categories. Viewed from this perspective, the raters in this study are expert attorneys who used three principle-based conceptual categories to cognitively represent numerous perceptions marked in the 15 performance dimensions. Recent researchers have argued that rating *halo errors* such as logical error might be more appropriately used as a measure of rater cognitive processing (Balzer & Sulsky, 1992). Guilford (1936) would have labeled the small number of categories as logical error, but a competing explanation is the effects of expert reasoning. A critical distinction in this study is that raters without legal expertise would <u>not</u> produce ratings on the 15 dimensions that could be concisely summarized by three factors. The influence of expert reasoning by raters can be inferred from a study by Van der Vleuten et. al. (1989). They found that rater training was least needed and least effective for expert raters (physicians) in comparison to novice raters (medical students) and lay raters who judged videotaped patient examinations using a detailed check-list.

The results of this study support the assertion that performance appraisals by expert attorney supervisors using 15 performance dimensions can be accurately summarized by one to three broad performance categories. The large first eigenvalues in Table 1 (11.8 and 11.5) indicate that much of the variance can be represented by a single *general impression halo* factor (Balzer & Sulsky, 1992). The eigenvalues for Factor 2 are much smaller (.96 and .88). The pattern of one large eigenvalue followed by a much smaller second value (11% to 20% of the first value) is also demonstrated by three factor analysis studies of medical student performance ratings (Dielman, Hull, & Davis, 1980; Forsythe, McGaghie & Friedman, 1985; Maxim & Dielman, 1987). This

pattern of one large eigenvalue followed by a second much smaller value suggests that expert raters in both medicine and law integrate their perceptions into a single emphatic composite judgment augmented by one or two uncorrelated judgments that are much less emphatic. Varimax rotations of the principal components produce factors that are easily interpreted.

Conclusions

Rater goal-directed PA theory (Murphy & Cleveland, 1995) provides an expanded framework for interpreting these research results in comparison to psycho physically oriented rating theory. In this study, the *rating context* is a civil service organization in which expert supervisors must produce "objective documentation" for employee promotion decisions. The *performance ratings* in this study are viewed as a communication from expert supervisors to both the organization and also to their supervisees. Ratings on a 15 dimension PA form are summarized by three broad performance categories. These three categories are labeled rater communication goals because the *rating context* requires summary *performance ratings* from supervisors.

This line of research could be extended by interviewing raters to better understand the relationship between their *performance judgment* (opinion of individual ratees) and the recorded *performance rating* (the written documentation). Moss (1996) provides a framework for integrating the *Naturalist* and *Interpretive* conceptions of social science. This type of integrated measurement research could help validate the four components of goal-directed rater PA theory.

Bibliography

Balzer, W.K., & Sulsky, L.M. (1992). Halo and performance appraisal research: A critical examination. *Journal of Applied Psychology, 77,* 975-985.

Cattell, R.B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1,* 245-276.

Murphy, K.R. & Cleveland, J.N. (1995). *Understanding Performance Appraisal: Social, Organization, and Goal-based Perspectives.* Sage Publications, Thousand Oaks, CA.

Dielman, T.E., Hull, A.L., & Davis, W.K. (1980). Psychometric properties of clinical performance ratings. *Evaluation & The Health Professions, 3,* 103-117.

Forsythe, G.B., McGaghie, W.C., & Friedman, C.P. (1985). Factor structure of the resident evaluation form. *Educational and Psychological Measurement, 45,* 259-264.

Glaser, R., & Chi, M.T.H. (1988). *The nature of expertise* (pp. xv-xxi). Hillsdale, NJ: Erlbaum.

Guilford, J.P. (1936). *Psychometric methods.* New York: McGraw-Hill.

Maxim, B.R., & Dielman, T.E. (1987). Dimensionality, internal consistency and interrater reliability of clinical performance ratings. *Medical Education, 21,* 130-137.

Messick, S. (1995). Validity of psychological assessment. *American Psychologist, 50,* 741-749.

Moss, P.A. (1996). Enlarging the dialogue in educational measurement: Voices from interpretive research traditions. *Educational Researcher, 25(1),* 20-27.

Ramsey PG Wenrich MD, Carline JD, Inui TS, Larson EB, & LoGerfo FP. (1993). Use of peer ratings to evaluate physician performance. *J. Am. Med. Assoc.,* V. 269(13), 1655-1660.

SAS Inc., (1989). *JMP User's Guide,* (pp. 451-471). Cary, NC: SAS Institute Inc.

Stevens, S.S., (1962). The surprising simplicity of sensory metrics. *American Psychologist, 17,* 29-39.

Van der Vleuten, C.P.M., Van Luyk, S.J., Van Ballegooijen A.M.J. & Swanson, D.B. (1989) Training and experience of examiners. *Medical Education, 23,* 290-296.